

(11) Publication Number: 2001-0017730

(43) Publication Date: March 5, 2001

(21) Patent Application Number: 10-1999-0033394

(22) Filing Date: August 13, 1999

(71) Applicant: Korea Eronic Telecommunication Research Institute

(72) Inventors: Hanmin JEON, Taewan KIM, Cheolmin SIM, Sungkwon CHOI, Sanghwa YEO, Youngkil KIM, Byeongrae RYU, Kwangjun SEO, Youngae SEO, Sanggyu PARK, Seyoung PARK, Yoonhyung ROH

(54) Title: AUTOMATIC TRANSLATOR USING SENTENCE-FRAME INCLUDING PROTECTOR AND SYNTAX NODE

**Abstract:**

The present invention relates to an automatic translator using sentence-frame including a protector and a syntax node, in which both context information and syntax information of an input sentence are utilized and a translated sentence with a natural meaning is generated.

Such an automatic translator using the sentence-frame including the protector and the syntax node includes: part-of-speech giving means for analyzing an inputted original text in morpheme units, giving a part of speech to each of fixedly expressed morphemes and giving a part of speech to each of remaining morphemes; sentence parsing means for marking important parts of speech in the original text to which the parts of speech are given with protectors, and attaching a proper syntax tag to partial syntaxes between the protectors; original frame search means for applying an original sentence-frame generated with the protectors and the partial syntaxes to an original sentence-frame database and searching a similar original sentence-frame of data; and translated text generating means for determining a proper translated sentence-frame among translated sentence-frames corresponding to the original sentence frame.

(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(51) Int. Cl.  
G06F 17/28

(45) 공고일자  
(11) 등록번호  
(24) 등록일자

2002년03월18일  
10-0329109  
2002년03월06일

(21) 출원번호	10-1999-0033394	(65) 공개번호	특2001-0017730
(22) 출원일자	1999년08월13일	(43) 공개일자	2001년03월05일
(73) 특허권자	한국전자통신연구원, 오길록 대한민국 305-350 대전 유성구 가정동 161번지		
(72) 발명자	정한민 대한민국 305333 대전광역시 유성구 어은동 1 김태완 대한민국 305333 대전광역시 유성구 어은동 99 한빛아파트 105-605 심철민 대한민국 302150 대전광역시 서구 만년동 초원아파트 110-602 최승권 대한민국 305333 대전광역시 유성구 어은동 99 한빛아파트 113-1101 여상화 대한민국 305335 대전광역시 유성구 궁동 395-3 다슬아파트 102-1105 김영길 대한민국 302222 대전광역시 서구 삼천동 보라아파트 202-1207 류병래 대한민국 132040 서울특별시 도봉구 창동 808 동아청솔아파트 114-1804 서광준 대한민국 305338 대전광역시 유성구 구성동 373-1 서영애 대한민국 305503 대전광역시 유성구 송강동 청송아파트 514-1109 박상규 대한민국 305333 대전광역시 유성구 어은동 99 한빛아파트 113-1505 박세영 대한민국 305333 대전광역시 유성구 어은동 한빛아파트 133-1004 노윤형 대한민국		

305350

대전광역시 유성구 가정동 한국전자통신연구원 기숙사 신관 225호

(74) 대리인

전영일

(77) 심사청구

심사관: 김재욱

(54) 출원명

프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기

## 요약

본 발명은 입력문장이 가지는 문맥정보와 구문정보를 모두 활용하면서 의미적으로 자연스러운 대역문을 만들어내는 프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기에 관한 것이다.

이러한 프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기는, 입력되는 원문을 형태소 단위로 분석한 후, 고정표현된 형태소들에게는 하나의 품사를 부여하고, 나머지 각 형태소들에게는 각각 하나의 품사를 부여하는 품사부여수단과; 상기 품사 부여된 원문 중에서 중요한 역할을 하는 품사들을 프로텍터들로 표시하고, 상기 프로텍터들 사이의 부분구문들에는 적절한 구문태그를 부착하는 문장분석수단; 상기 프로텍터들과 부분구문들로 생성된 원문틀을 원문틀 데이터베이스에 적용하여 유사한 데이터 원문틀을 찾는 원문틀 검색수단; 및 상기 원문틀에 대응하는 대역문틀들 중 적절한 하나의 대역문틀을 결정하여 대역문을 생성하는 대역문 생성수단을 구비한다.

## 대표도

도1

명세서

## 도면의 간단한 설명

도 1은 본 발명에 따른 문틀을 이용한 자동번역기를 도시한 구성 블록도.

도 2는 본 발명에 따른 문틀을 이용한 자동번역기의 영한번역의 한 실시예를 도시한 도면이다.

## ※ 도면의 주요부분에 대한 기호의 설명 ※

101 : 입력부 102 : 형태소 분석부

103 : 품사 결정부 104 : 고정표현 인식부

105 : 프로텍터 발견부 106 : 부분구문 분석부

107 : 원문틀 생성부 108 : 원문틀 발견부

109 : 대역문틀 선택부 110 : 대역어 생성부

111 : 인쇄제어부 112 : 인쇄장치

113 : 표시제어부 114 : 표시장치

## 발명의 상세한 설명

## 발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 프로텍터와 문틀을 이용하여 원문으로부터 자연스러운 대역문을 만들어내는 기계 번역에 관한 것으로서, 보다 상세하게 설명하면 입력문장이 가지는 문맥정보와 구문정보를 모두 활용하면서 의미적으로 자연스러운 대역문을 만들어내는 자동번역기에 관한 것이다.

기계 번역은 1950년대부터 연구되기 시작하면서 다양한 번역 기법을 선보여 왔다. 이 기계 번역의 가장 보편적인 방법으로서 원문에 대해 어휘 분석과, 구문 분석, 변환, 및 생성단계들을 통해 대역문을 생성하는 방법이 있다. 이에 관한 선행특허로서, 미국특허 제5826219호(등록일 1998년 10월 20일, 권리권자 Kutsumi, 발명의 명칭 'Machine translation apparatus')가 있다. 이는 직관적이며 간단하기는 하나, 실생활에 나타나는 언어 현상을 파악하지 못하고 문법적으로만 판단한 모든 가능한 조합의 결과를 생성해냄으로써 수많은 애매성을 발생시키고 의미적으로 이해할 수 없는 대역문들을 만들어낸다는 문제점을 가진다.

이의 극복을 위해 위의 방법에 속어 인식을 추가하거나 구문 필터링/선택 기법 등을 도입하였으나, 이러한 방법도 가능한 모든 조합의 구문 분석을 수행함으로써 올바른 대역문장을 생성하지 못하였다.

또 다른 방법으로서, 권리권자가 Maruyama이고, 발명의 명칭이 'Method and apparatus of translation based on patterns'인 미국특허 제5640575호(등록일 1997년 6월 17일)에는, 원시언어의 문장 예제(문장패턴)와 일대일로 대응하는 대역 문장패턴을 메모리에 저장한 후, 입력문장과 일치하는 원시문장패턴을 찾아서 그에 대응하는 대역문장패턴을 출력하는 번역방법이 개시되어 있다.

이러한 기계번역방법은 이미 구축되어 있는 패턴들에 대해서는 어느 정도의 성능을 보장하였으나, 현실적으로 이러한 문장패턴들을 모두 구축하는 것이 불가능하다. 따라서, 이러한 기계번역방법은 제한된 영역에서만 실험적으로 사용되어 왔다.

상기한 번역방법 외에도 통계학적 기반의 번역방법이 있으나, 이는 학습된 분야에서만 일정한 성능을 발휘하며, 그 학습 영역을 넓히기 힘들다는 문제점은 여전히 해결하지 못하였다.

#### 발명이 이루고자 하는 기술적 과제

따라서, 본 발명은 상기와 같은 종래기술의 문제점을 해결하기 위하여 안출된 것으로서, 본 발명의 목적은 기존의 단일 접근적인 방법론을 벗어나 복합적인 번역 개념을 프로텍터와 문틀에 반영하여 고품질의 번역이 가능한 프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기를 제공하는 데 있다.

#### 발명의 구성 및 작용

상기한 목적을 달성하기 위한 본 발명에 따르면, 프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기가 제공된다. 이는 입력되는 원문을 형태소 단위로 분석하는 형태소 분석수단; 상기 형태소 단위로 분석된 각각의 단어들에 품사를 결정하는 품사 결정수단; 상기 품사 결정수단에서 품사가 결정된 단어들 중 하나의 단어나 구처럼 취급되는 속어, 복합 명사, 연어 등의 고정표현된 단어들을 묶어서 하나의 품사를 재결정하여 부착하는 고정표현 인식수단; 상기 품사 결정수단과 고정표현 인식수단에서 부여된 품사 중에서 문장에서 중요한 역할을 하는 품사들(동사, 접속사, 관계사, 기호)을 프로텍터로 표시하는 프로텍터 발견수단; 상기 프로텍터 발견수단에서 발견된 프로텍터들 사이의 부분구문이 명사구 또는 전치사구인 지를 분석하여 해당 구문태그를 부착하는 부분구문 분석수단; 상기 프로텍터 발견수단에서 발견된 프로텍터들과 상기 부분구문 분석수단에서 부착된 부분구문들로 이루어진 원문의 문틀을 원문을 데이터베이스에 적용하여 유사한 데이터 원문틀을 찾는 원문틀 검색수단; 및 상기 원문틀 검색수단에서 찾아진 데이터 원문틀에 대응하며 대역언어의 구조적 특성을 모두 포함한 번역 구조의 대역문틀들을 찾아서 대역문을 생성하는 대역문 생성수단을 구비한 것을 특징으로 하는 프로텍터와 구문노드를 포함하는 것을 특징으로 한다.

또한, 본 발명에 따르면 위에서 설명하였던 프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기를 기능시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체가 제공된다.

이하, 첨부된 도면을 참조하여 본 발명의 한 실시예에 따른 '프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기'를 보다 상세하게 설명하면 다음과 같다.

본 실시예에서 프로텍터(protector)라 함은, 동사, 접속사 등과 같이 문장에서 구문 분석의 애매성이 폭발적으로 증가하기 시작하는, 문장의 특성을 반영한 단어를 의미한다. 프로텍터를 인식하면 문장의 형태를 파악할 수 있으며, 번역을 위한 단서를 얻을 수 있다.

또한, 본 실시예에서 문틀이라 함은, 문장이 가공되어 단순화된 형태를 말하며, 문틀의 구성 요소를 슬롯이라 하는데, 슬롯은 프로텍터 또는 프로텍터 사이의 구문노드로 이루어진다. 여기서, 구문노드란 프로텍터 사이의 문장 일부를 구문 분석하여 얻은 구문적 성격을 포함하는 것으로, 명사구는 NP, 전치사구는 PP와 같이 표현된다.

도 1은 본 발명의 한 실시예에 따른 프로텍터와 구문노드를 포함하는 문틀을 이용한 자동번역기를 도시한 구성 블록도이다. 이는 입력부(101)와, 형태소 분석부(102), 품사 결정부(103), 고정표현 인식부(104), 프로텍터 발견부(105), 부분구문 분석부(106), 원문을 생성부(107), 원문을 발견부(108), 대역문을 선택부(109), 대역어 생성부(110), 인쇄제어부(111), 인쇄장치(112), 표시제어부(113), 및 표시장치(114)로 구성된다.

형태소 분석부(102)에는 번역사전이 제공되며, 대역어 생성부(110)에는 번역사전과 대역어 생성규칙이 제공된다. 또한, 고정표현 인식부(104)에는 고정표현 데이터베이스가 제공되고, 원문을 발견부(108)에는 원문을 데이터베이스가 제공된다. 원문을 데이터베이스에는 임의의 데이터 원문들에 대응하는 여러 개의 데이터 대역문들이 저장된다.

원문은 입력부(101)를 통해 들어온다. 형태소 분석부(102)는 입력된 원문을 각각의 단어에 대해 형태소를 분석하고, 품사 결정부(103)는 분석된 각각의 단어들에 품사를 결정한다. 고정표현 인식부(104)는 품사 결정부(103)에서 결정된 품사를 재결정하여 부착하는데, 속어, 복합 명사, 연어 등과 같이 하나의 단어나 구처럼 취급되는 단어들을 묶어서 그에 해당하는 새로운 품사를 부착한다. 이와 같이 고정표현을 인식함으로써 구문분석이나 문들의 설정을 보다 단순화시킬 수 있다.

프로텍터 발견부(105)는 문장에서 중요한 역할을 하는 품사나 단어들, 예를 들어 동사, 접속사, 관계사, 기호 등을 발견하여 프로텍터라는 표시를 부착한다. 이때, 이들 프로텍터들에 대해서는 어떠한 분석도 수행하지 않는다. 부분구문 분석부(106)는 프로텍터들 사이의 단어들에 대해 부분구문 분석을 실시하여, 그들에 대해 적절한 구문 태그들을 부착한다. 프로텍터를 제외한 부분구문 분석이므로 애매성이 거의 발생하지 않으며, 분석 시간도 상당히 단축된다.

원문을 생성부(107)는 앞에서 결정된 프로텍터들과 구문 태그들을 이용하여 입력된 원문에 대한 원문들을 만들어낸다. 원문을 발견부(108)는 이렇게 만들어진 입력된 원문에 대한 원문들을 입력받아서 원문을 데이터베이스 내에서 저장된 데이터 원문들과 비교하면서 동일한 데이터 원문들을 발견한다. 원문에 대한 원문들과 일치하는 데이터 원문들이 없는 경우에는 여러 데이터 원문들을 한꺼번에 이용하여 원문을 커버하도록 한다. 이렇게 발견된 데이터 원문들은 여러 개의 데이터 대역문들을 가지고 있으며, 이 대역문들은 이미 대역언어의 구조적 특성을 모두 포함하고 있는, 현실적으로 가능한 번역 구조만을 포함하고 있다.

대역문 선택부(109)는 상기와 같이 찾아진 데이터 대역문들 중에서 문맥에 따라 하나의 대역문을 선택한다. 즉, 공기 정보, 구문/의미 정보들을 이용하여 현재의 문맥에 맞는 하나의 대역문을 선택한다. 대역어 생성부(110)는 이렇게 선택된 대역문들로부터 그 대역문들이 가지는 각각의 슬롯을 구조 변환하여 완전한 하나의 문장으로 만들어낸다.

이와 같이 입력된 원문에 대응하도록 만들어진 대역문은 인쇄제어부(111)를 통해 인쇄장치(112)에 프린트되거나, 표시제어부(113)를 통해 표시장치(114)에 디스플레이된다.

도 2는 본 발명의 한 실시예에 따른 문들을 이용한 자동번역방법을 순차적으로 설명하기 위해 도시한 영한 번역의 한 실시예이다.

도 2의 (a)에는 원문으로서 영어 입력 문장이 'The White House said the president decided to grant duty-free status for 18 categories, but turned down such treatment for other types'로 주어진다.

이 원문에 대해서 형태소 분석단계 및 품사 결정단계를 거치면, 도 2의 (b)와 같은 결과가 출력된다. 'The(det) White(noun) House(noun) said(verb) the(det) president(noun) decided(verb) to(preposition) grant(verb) duty-free(noun) status(noun) for(preposition) 18(noun) categories(noun) ,(symbol) but(conjunction) turned(verb) down(adv) such(adj) treatment(noun) for(preposition) other(adj) types(noun)'

여기서, det는 정관사, noun은 명사, verb는 동사, prep는 전치사, symbol은 기호, conj는 접속사, adv는 부사, adj는 형용사를 각각 지칭한다.

이와 같이 각 단어들의 형태소가 분석되고 품사가 결정된 상태에서, 속어, 복합명사, 연어 등과 같이 하나의 단어나 구처럼 취급되는 단어들을 하나로 묶는 고정표현 인식단계를 수행한다. 그 결과는 도 2의 (c)와 같다. 즉, 'The(det)

White House  
(noun) said(verb) the(det) president(noun)

decided to grant  
(verb) duty-free(noun) status(noun) for(preposition) 18(noun) categories(noun) ,(symbol) but(conj)

turned down  
(verb) such(adj) treatment(noun) for(preposition) other(adj) types(noun)'

즉, 상기한 고정표현 인식단계를 거치면, 복합명사의 2개의 명사들 'White'와 'House'가 하나의 명사 'White House'로 인식되고, 속어인 하나의 동사 'turned'와 하나의 부사 'down'가 하나의 동사 'turned down'로 인식된다. 또한, 연어인 세 단어들 'decided(verb)'와 'to(preposition)'와 'grant(verb)'가 하나의 동사 'decided to grant'로 인식된다.

이와 같이 고정표현을 인식한 상태에서, 문장에서 중요한 역할을 하는 품사나 단어들, 예를 들어 동사, 접속사, 관계사, 기호등의 프로텍터를 발견한다. 그 결과는 도 2의 (d)와 같다. 즉, 'The(det) White House(noun)

said

(verb) the(det) president(noun)

decided to grant

(verb) duty-free(noun) status(noun) for(pre) 18(noun) categories(noun)

(symbol/comma)

but

(conj/but)

turned down

(verb) such(adj) treatment(noun) for(pre) other(adj) types(noun)'

이와 같이 프로텍터들을 찾은 후에, 찾아진 프로텍터들 사이의 단어들에 대해 부분구문 분석을 실시한다. 즉, 프로텍터들 사이의 단어들을 명사구와 전치사구로 구분하는데, 그 결과는 도 2의 (e)와 같다. 'NP ← The White House, NP ← the president, NP ← duty-free status for 18 categories, NP ← such treatment, PP ← for other types' 분석된 부분구문들에는 적절한 구문태그(NP, PP)가 부착된다.

이와 같이 부분구문을 분석한 후, 이들 적절하게 부착된 구문태그들과 앞에서 발견된 프로텍터들을 이용하여 원문들을 생성한다. 생성된 원문들은 도 2의 (f)와 같다. 즉, 'NP1 verb1 NP2 verb2 NP3 symbol/comma conj/but verb3 NP4 PP/for' 이와같이 입력문장(원문)에 대한 원문들이 생성되면, 이를 원문을 데이터베이스 내에서 동일한 데이터 원문들을 발견하는 데 사용한다.

동일한 데이터 원문들이 발견되면, 이 데이터 원문들에 대응하는 대역문들 중에서 가장 적절한 대역문들을 선택한다. 선택된 대역문들은 도 2의 (g)와 같다. 즉, 'NP1(은) NP2(이) NP3(을) verb2(다고) verb1 symbol/comma conj/but PP NP4(는) verb3'

이와 같이 대역문들이 선택되면, 그 대역문들이 가지는 각각의 슬롯들을 구조 변환하여 하나의 완전한 문장을 만들어 낼 수 있다. 생성된 대역문은 도 2의 (h)와 같다. 즉, '백악관은 대통령이 18개의 항목을 위한 세금면제상태를 인정하기로 결정했다고 말했다. 그러나 다른 종류에 대한 그러한 처리는 거절했다'

이와 같이 본 발명에 따르면, 입력 영어 문장이 주어지면, 형태소 분석 및 품사 결정단계와, 고정표현 인식단계, 프로텍터 발견단계, 부분구문 분석단계, 원문을 생성 및 발견단계, 대역문을 선택단계, 및 대역문 생성단계를 통해 완전하고 정확하게 번역된 한국어 문장을 생성할 수 있다.

위에서 양호한 실시예에 근거하여 이 발명을 설명하였지만, 이러한 실시예는 이 발명을 제한하려는 것이 아니라 예시하려는 것이다. 이 발명이 속하는 분야의 숙련자에게는 이 발명의 기술사상을 벗어남이 없이 위 실시예에 대한 다양한 변화나 변경 또는 조절이 가능함이 자명할 것이다. 그러므로, 이 발명의 보호범위는 첨부된 청구범위에 의해서만 한정될 것이며, 위와 같은 변화예나 변경예 또는 조절예를 모두 포함하는 것으로 해석되어야 할 것이다.

#### 발명의 효과

이상과 같이 본 발명에 의하면, 모든 분야에 있어서 원문으로부터 고품질의 자연스러운 대역문을 고속으로 만들어 낼 수 있기 때문에, 자연어 처리의 전 분야에서 번역 기능을 제공할 수 있도록 한다.

#### (57) 청구의 범위

##### 청구항 1.

원시언어로 입력되는 원문을 형태소 단위로 분석하는 형태소 분석수단;

상기 형태소 단위로 분석된 각각의 단어들에 품사를 결정하는 품사 결정수단;

상기 품사 결정수단에서 품사가 결정된 단어들 중 하나의 단어나 구처럼 취급되는 속어, 복합 명사, 연어 등의 고정표현된 단어들을 묶어서 하나의 품사를 재결정하여 부착하는 고정표현 인식수단;

상기 품사 결정수단과 고정표현 인식수단에서 부여된 품사 중에서 문장에서 중요한 역할을 하는 품사들(동사, 접속사, 관계사, 기호)을 프로텍터로 표시하는 프로텍터 발견수단;

상기 프로텍터 발견수단에서 발견된 프로텍터들 사이의 부분구문이 명사구 또는 전치사구인 지를 분석하여 해당 구문태그를 부착하는 부분구문 분석수단;

상기 프로텍터 발견수단에서 발견된 프로텍터들과 상기 부분구문 분석수단에서 부착된 부분구문들로 이루어진 원문의 문들을 원문을 데이터베이스에 적용하여 유사한 데이터 원문들을 찾는 원문을 검색수단; 및

상기 원문을 검색수단에서 찾아진 데이터 원문들에 대응하며 대역언어의 구조적 특성을 모두 포함한 번역 구조의 대역문들들을 찾아서 대역문을 생성하는 대역문 생성수단을 구비한 것을 특징으로 하는 프로텍터와 구문노드를 포함하는 문들을 이용한 자동번역기.

청구항 2.  
청구항2는 삭제 되었습니다.

청구항 3.  
청구항3는 삭제 되었습니다.

청구항 4.  
청구항4는 삭제 되었습니다.

청구항 5.  
제1항에 있어서, 상기 원문을 검색수단은, 입력된 원문의 원문들과 정확하게 일치하는 데이터 원문들이 없는 경우에는 여러 개의 데이터 원문들을 조합하여 사용하는 것을 특징으로 하는 프로텍터와 구문노드를 포함하는 문들을 이용한 자동번역기.

청구항 6.  
제1항에 있어서, 상기 대역문 생성수단은, 상기 데이터 원문들에 대응하는 여러 개의 대역문들 중 프로텍터의 타입, 공기 정보, 의미정보, 형태들을 이용하여 하나의 대역문들을 선택하는 것을 특징으로 하는 프로텍터와 구문노드를 포함하는 문들을 이용한 자동번역기.

청구항 7.  
제1항에 있어서, 상기 대역문 생성수단은, 대역문들이 가지는 각 구문태그들과 프로텍터들에 대해 대역구문구조를 생성하고, 각 단어들의 대역어를 결정하여 전체적인 대역문을 생성하는 것을 특징으로 하는 프로텍터와 구문노드를 포함하는 문들을 이용한 자동번역기.

청구항 8.  
컴퓨터에,

입력되는 원문을 형태소 단위로 분석하는 형태소 분석수단;

상기 형태소 단위로 분석된 각각의 단어들에 품사를 결정하는 품사 결정수단;

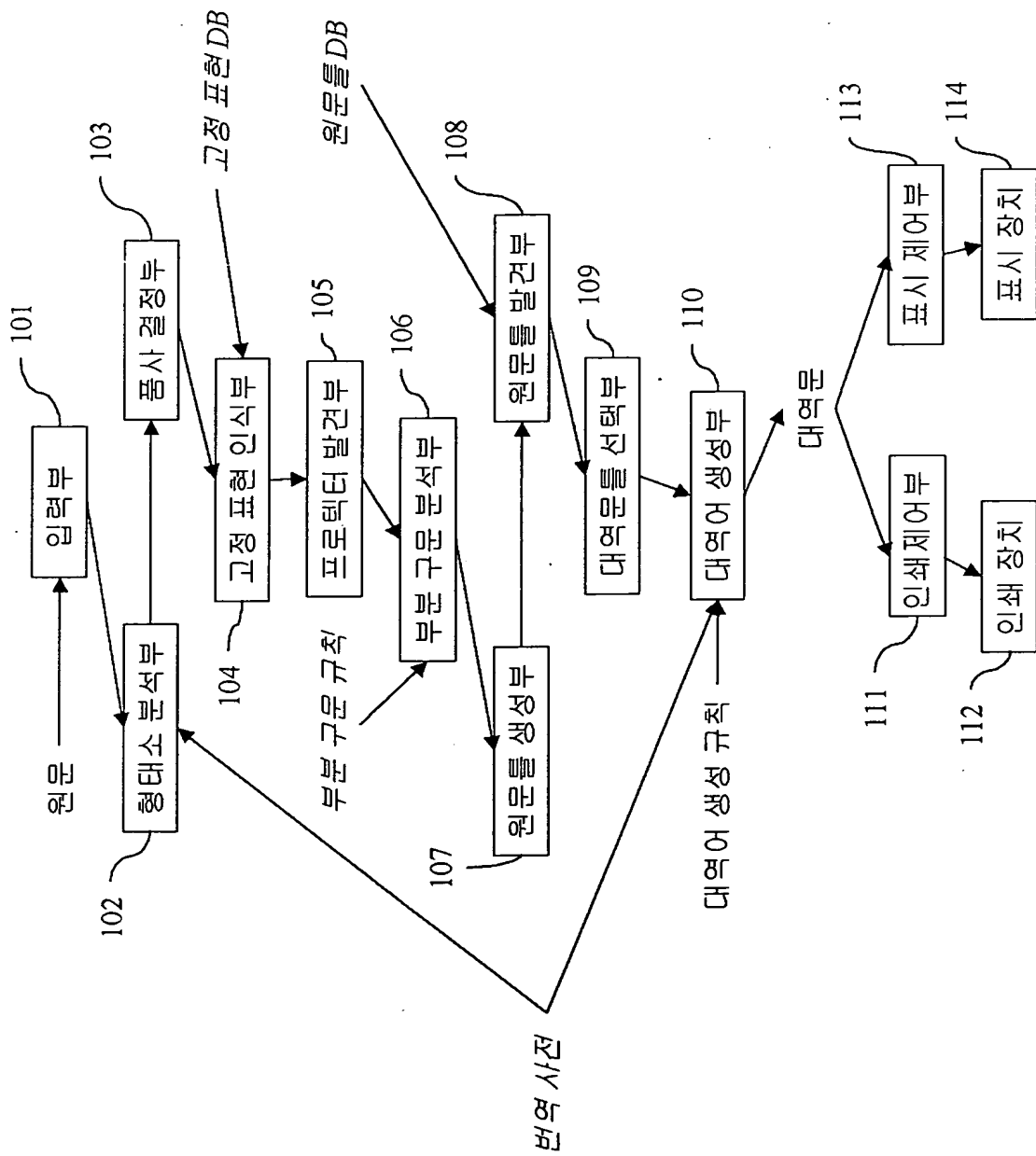
상기 품사 결정수단에서 품사가 결정된 단어들 중 하나의 단어나 구처럼 취급되는 속어, 복합 명사, 연어 등의 고정표현된 단어들을 묶어서 하나의 품사를 재결정하여 부착하는 고정표현 인식수단;

상기 품사 결정수단과 고정표현 인식수단에서 부여된 품사 중에서 문장에서 중요한 역할을 하는 품사들(동사, 접속사, 관계사, 기호)을 프로텍터로 표시하는 프로텍터 발견수단;

상기 프로텍터 발견수단에서 발견된 프로텍터들 사이의 부분구문이 명사구 또는 전치사구인 지를 분석하여 해당 구문태그를 부착하는 부분구문 분석수단;

상기 프로텍터 발견수단에서 발견된 프로텍터들과 상기 부분구문 분석수단에서 부착된 부분구문들로 이루어진 원문의 문들을 원문을 데이터베이스에 적용하여 유사한 데이터 원문들을 찾는 원문을 검색수단; 및

상기 원문들에 대응하는 대역문들을 찾아서 대역문을 생성하는 대역문 생성수단을 기능시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.





## [영한 번역]

- (a) 입력 문장: The White House said the president decided to grant duty-free status for 18 categories, but turned down such treatment for other types
- (b) 형태소 분석 및 품사 결정: The(det) White(noun) House(noun) said(verb) the(det) president(noun) decided(verb) to(preposition) grant(verb) duty-free(noun) status(noun) for(preposition) 18(noun) categories(noun), (symbol) but(conjunction) turned(verb) down(adv) such(adj) treatment(noun) for(preposition) other(adj) types(noun)
- (c) 고정 표현 인식: The(det) White House(noun) said(verb) the(det) president(noun) decided to grant(verb) duty-free(noun) status(noun) for(preposition) 18(noun) categories(noun), (symbol) but(conjunction) turned down(verb) such(adj) treatment(noun) for(preposition) other(adj) types(noun)
- (d) 프로텍터 발견: The(det) White House(noun) said(verb) the(det) president(noun) decided to grant(verb) duty-free(noun) status(noun) for(preposition) 18(noun) categories(noun), (symbol/comma) but(conj/but) turned down(verb) such(adj) treatment(noun) for(preposition) other(adj) types(noun)
- (e) 부분 구문 분석: NP ← The White House, NP ← the president, NP ← duty-free status for 18 categories, NP ← such treatment, PP ← for other types
- (f) 원문을 생성 및 발견: NP1 verb1 NP2 verb2 NP3 symbol/comma conj/but verb3 NP4 PP/for
- (g) 대역문을 선택: NP1(은) NP2(이) NP3(을) verb2(다고) verb1 symbol/comma conj/but PP NP4(는) verb3
- (h) 대역어 생성: 백악관은 대통령이 18개의 항목을 위한 세금 면제 상태를 인정하기로 결정했다고 말했다, 그러나 다른 종류에 대한 그러한 처리는 거절했다